

22 107.4 5.0 259498 61 AC020876
23 104.8 4.9 731 77 CNSOIFEV
24 100.4 4.6 11886 1 AE004638
25 94.8 4.4 110000 71 AC078913_1
26 87.6 4.1 41807 29 SC6011
27 84.6 3.9 16164 1 AE004094
28 72.6 3.4 1713 81 A07753
29 71.8 3.3 15637 1 AE001730
30 69.8 3.2 96086 29 STYSTMD1
31 67.4 3.1 231916 73 AL158037
32 66.4 3.1 146174 2 D90910
33 64.2 3.0 13725 1 AE000832
34 62.4 2.9 1788 1 AF091510
35 62.4 2.9 7045 2 ECPROM
36 62.4 2.9 7203 2 E01LVGMED
37 62.4 2.9 8088 2 E01LVGMED
38 62.4 2.9 14325 1 AE000453
39 62.4 2.9 91414 2 E00W85
40 60.8 2.8 2343 2 E01LVX
41 60.8 2.8 2841 21 E10718
42 60.8 2.8 2841 81 AR092608
43 60.8 2.8 2841 81 E13058
44 60.8 2.8 4900 2 ECORNLIV
45 60.8 2.8 9456 2 E00ILVGE

ALIGNMENTS

RESULT 1
SCIA9 29400 bp DNA BCT 11-DEC-1998
LOCUS Streptomyces coelicolor cosmid 1A9.
AC034446
VERSION 1 GI:4007685

KEYWORDS
ADA-like regulatory protein; arac family; gntR; helix-turn-helix;
His rich; integral membrane; luxR; luxR family;
methylated-DNA-protein-cysteine methyl transferase; nucleotide
binding; ogt; oxidoreductase; poxB; pyruvate oxidase; response
regulator; sensor kinase; sorbitol oxidase; thiamine
pyrophosphate; TPP; transferase; transmembrane protein; two
component system.

SOURCE
Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).

REFERENCE
AUTHORS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
JOURNAL Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 29400)
AUTHORS Saunders,D.C. and Harris,D.
JOURNAL Unpublished
TITLE 2 (bases 1 to 29400)
Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
JOURNAL Direct Submission

REFERENCE
AUTHORS Submitted (04-DEC-1998) Streptomyces coelicolor sequencing project,
JOURNAL Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
REFERENCE CB10 ISA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
AUTHORS David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
JOURNAL Colney, Norwich, Norfolk NR4 7UH, UK
TITLE 3 (bases 1 to 29400)
Redenbach,M., Kieser,H.M., Denapante,D., Eichner,A., Cullum,J.,
JOURNAL Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
JOURNAL for the 8 Mb Streptomyces coelicolor A3(2) chromosome
MEDLINE Mol. Microbiol. 21 (1), 77-96 (1996)
COMMENT 97000351

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S-coelicolor/>) CDS are
numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).

The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or att) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid 1A9 lies
between 9B2 and 6C5 in the Ase-I-B genomic restriction fragment.

FEATURES
Location/Qualifiers
1..29400

/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"

/clone="cosmid 1A9"
complement(1..651)
/gene="SC1A9.01c"

complement(<1..651)
/gene="SC1A9.01c"

/note="SC1A9.01c, incomplete CDS, possible transferase,
partial CDS, len: 219 aa, similar to TR:O53185
(EMBL:AL021246) a proposed transferase from Mycobacterium
tuberculosis (302 aa), fasta scores: Opt: 764, z-score:
1097.3, E(): 0, (62.1% identity in 203 aa overlap (302
aa). Also weakly similar to several methyltransferases eg.
METH_ECOLI
(EMBL:X16584)5-methyltetrahydrofolate-homocysteine
methyltransferase (1226 aa), fasta scores: Opt:160,
z-score: 234.8, E(): 8.4e-06, (30.0% identity in 220 aa
overlap)."

/codon_start=1
/transl_table=11
/product="putative transferase"

/protein_id="CAA22371.1"
/db_xref="GI:4007686"

/translation="MTSDFADALASGLVLDGSLNQLAAGHLDGALSARLLAED
PEAITRAHLAYFEAGAEVAITSSYQATFEFGARRGIGRERAAELLSVASAREARR
ARTAPRALWVAASGAYGAMLDGSEYGRYGLGRGALERHPRLEVLAAAPDV
LALETVPDTEAAALLRAVRGLDVPALSYTVAGDTRAGQPLDEAFALAADVDEVI"
731..1654
/gene="SC1A9.02"

/note="SC1A9.02, possible transmembrane protein, len:
307aa; Contains several possible membrane spanning
domains."

/codon_start=1
/transl_table=11
/product="putative transmembrane protein"

/protein_id="CAA22372.1"
/db_xref="GI:4007687"

/translation="MTISGRRTKRSVRCSPRSHSGHDEGVVPVRFVWQFLAVLVAIAIG
GIAQVAKNDMLTLVGLTVALVYVAVVVRTERREALDGLDVAAGAGMTL
IGFLFGAVITNLFASGYVDGGLGALVGPMAAAATEEVFGVLFRIEE
HIGTYLALGLTGLVFCGLMHLNEDATLWALALAIAGFWLAAAYATRNMLTGVH
FGWNPAGGVSTVSGNGDSGLLDATMSGPKLLTGGDFGPGGSYISVGFGLVLTLY
FWLHRRGNIVAFGSRRAAGANSAAITLPR"
1651..2817
/gene="SC1A9.03"

gene

CDS

gene

○ 〇

CD3

```
/transl table=11
```

```

/product="putative transmembrane protein"
/protein_id="CAA22375.1"

```

```

/db_xref="GI:4007690"
/translation="MFRNAGIRAPYVHVPMRGDMASRSDDRPRLRRDRLSGWIGYMALGAGVYRVLNGTTSVALLCAVAAVYVYVWVRRVRDARAVGAEPPDYPAMEROIILKGGPQDPERRAMAFVDSROELRRHPHMFPLAVIFGT"SAWMTLSGSGVSGAGSLMLGLGVFLGNLWYINLDRLSHNRGLRG"
4140..4631
/gene="SCL1A9.06"
4140..4631
/gene="SCL1A9.06"

```

	Query Match	24.68;	Score 532.4;	DB 29;	Length 29400;
	Best Local Similarity	58.0%;	Pred. No. 9.8e-131;		
	Matches 1020;	Conservative 0;	Mismatches 726;	Indels 12;	Gaps 4;
332	acacagctaccgaacaataatgatcacatttggaaagtccaaggttgaaagcgaattta 391				
18101	ACAGAACGTCGGGAACAAGTTCTGCACATCCTCACCCGGCGGGGTGCAAGGCCCTTA 18160				
392	tggtttgtggtgacagccctaactcgatcgtagtgcgtgt---cgcgcaatcagatat 448				
18161	CGGAGTCTGTGGGACACAGCCTCAACCCCGTCTGTGGACGCCGTGCGCCGCCCATTCGCGCAT 18220				
419	tgagtgggtgcacgttcogaaatgaggaaoggcgccgttttgcagccggtgcgaatactgt 508				
18221	CGAATTGGTGACGTATACGGCAGGAGACACCGCGCCTTCGCCCGCGGGCGGAGGGCCA 18280				
509	gatacactgggagctggcagtatgtctgtcttcttgggtctcttgaaacacacacctgat 568				
18281	GATCACCCGGGAAGCTCACCGGTGCGCGGCTCTGTGCGCCCGCGCAACTCCACCTCAT 18340				
569	tcagggtctttatgatctgcatacaaatggctgcaagggtgttgcccatcgcctagccatat 628				
18341	CAACGGGTCTTACGACGCCACCGCTTCATGGCCCCCGTCTCGCCCTGGCCTCCCAGAT 18400				
629	tccagtgcccagattdggttcgaagcttcttcaggaaaacgcatcgcgagattttgttta 688				
18401	CCCGTCCAGCAGATCGGGCTCGGCTTCTTCCAGGAGACCACCCCGACACGCTGTTCCG 18460				
689	ggaatgcttggttacttcgagaatgggtgaatggtgtgtgagcaggggtgaacgcattttgca 748				
18461	CGAGTGCAGTCACTACAGCAGAGCTGATCTCCAGCCCGAAAGCAGATGCGCCGGTGTCTCCA 18520				
749	tcacgcgatctagtcaccatggcgggttaaagggtgtctcgtagtagtgcctcgtgga 808				
18521	GACCGCCATCCACACGCCGTGGCCACGGCGGGCGGTGAGGGTGTCTTCGTGCCCCGCCGA 18580				
809	tatcgtaaggaaagcgcaggtgacgttacatttccaaattccactattctctctggcac 868				
18581	CATCGCGACGAGCCCGCCCGCAGGGGGCGCGAGACGGGCCCTCGTCACTTCCGGCC 18640				
869	tecttggtgttcocggatactactgaggttcgagcgtggttgagggcgattacaacgc 928				
18641	CACCGTCCGG---CCCGCGCAGAGAGATCGACCGCGCTGCTCGGATGATCGACACGC 18697				
929	taagtctgtaacttgtcttcgctgcggcggtgaagaatgcicgcgcgcaggtgttga 988				
18698	CGAAGGTCAACCTGTCTCGGCACGGCACCCCGCGGGCGGACGCCGAGGTCAATGGA 18757				
989	gttgccggagaagattaaatcaccgatcgggcatgctgctgggttggttaagcaglacatcca 1048				
18758	GTTCCCGGGGAAGTCAAGGACCCGGTCTGGCACGCCCTTCGGGGGCAAGGAGTTTCATCCA 18817				
1049	gcatagaataccgtttgaaggtcggcatagtcttgccctgtcttggttaagcgccctgcgtga 1108				
18818	GTACGCACAACCCGTACACAGTTCGCATGACGGAGTCTCTTCATCGGCACCGACTTCCCGTACACGCCCTT 18937				
18878	GGCCACCCACGATGCCAGACTGCTCTCTTCATCGGCACCGACTTCCCGTACACGCCCTT 18937				
1109	tgcttccaatgagcggatctgctgattctcttatitgggtaacggatttcccttatctgattt 1168				

BEST AVAILABLE COPY

Qy 1169 ccttctaaagac---aacgttcccaggctgagatatacaacggtgagcacattggtgagc 1225
Db 18938 CCTCCGGACACGCTGAAGATCGCCAGATCGAGCTGCGGCCCGACGACCTGGGACGGCG 18997
Qy 1226 tacacaggtgaagatattccgtgaccgtggtggtgctgcgaacaatgaaatattttgccc 1285
Db 18998 CTCACAGCTGGACCTCCGGTGTGGGCGACGCGCGGAGACGCTGCGCTGCCTGATCCC 19057
Qy 1286 tcattgaagaaacacagatgcttcccttctgagtcggtatcgaagcacaacagcgcg 1345
Db 19058 CCGGCTCAAGAGAGAAAGAACCCGCTTCTTGGACCGGATGCTGAAGAGACGCGCGA 19117
Qy 1346 taagttagctcgggtgtagaagatcacacataacgctcgagagacgatgctcatttca 1405
Db 19118 CGCCCTGAAGCGCTGCTCAAGGGGTACACCCGCAAGGTGACACAGACGCTGCCGATCCA 19177
Qy 1406 cccgaatacattccctctattttgaacagctggtgaggaatgagtgctgtttactgt 1465
Db 19178 CCGCGAGTACGTGGCGGCCCTCTGCTGGACGAGATGGCCGACGACGCGGTTTCACCGT 19237
Qy 1466 gataccggcatgtgcaatgtgtgcatgctgaggtacatcgagaatccggagggaacgcg 1525
Db 19238 CGACACCGGATGTCACAGCTGTGGCGGCTCGTTACATCTCG---CCCAACGGCGCGC 19294
Qy 1526 cgacttctggttctattccgcacgagctggtgaggaatgagtgctgtttactgt 1585
Db 19295 CCGCATCATCGTCTCTTCTCCCGGCTTCGATGCGCAACGCGCTGCCGATGCCGATCGG 19354
Qy 1586 tgcgcaaggtgtgtagcaacacgacgagtgatcgatgctggtgaggaatgagtgctgt 1645
Db 19355 CGCCCAAGTTACCGACCGCGCGGCTGCTGCGATGCTGCGGACGCGCGGCTTCAC 19414
Qy 1646 catgctgctgggtgagcttctgacgttaagtgcacaaactccgctgaagcgtgagc 1705
Db 19415 CATGCTGATGGCGGACTTCTTCCACCTCTCCAGCAGCAGCTGCCGCTGAAGATCGTCT 19474
Qy 1706 gtttaacacagttcttctggtgagtgagtgagtgagtgagtgagtgagtgagtgag 1765
Db 19475 CTTCAACAACTCTCGCTCGGCAATGCTGAGTGGAGATGCTGCTGCGGCGGCTGCCCTC 19534
Qy 1766 attgttactgacatgaggaagtgaattcgcagagattggtggtggtggtggtggtggt 1825
Db 19535 GCACGGGGTGGCCAAAGAACCCGAGCTTCCGCGCGCTGCGCGAGGCTGCGCGGCTT 19594
Qy 1826 atcggtacgcatcacgataccgagaaagtctgcgagcagctgctgagcattgagcata 1885
Db 19595 CGCGCTCCGGTGGAGAAGCCCAAGGACCTGCGCGGCGCTGAAGGCGGCTTCAAGCA 19654
Qy 1886 tcttgacctgactgactgactgactgactgactgactgactgactgactgactgactgact 1945
Db 19655 CAAGGGTCCGCGCTCTGCTGACGCTGCTGACGACCCCAAGCGCTGCTGCTGCTGCTG 19714
Qy 1946 catcacgtggaaacagctcagtgatgagtgatgagtgatgagtgatgagtgatgagtgat 2005
Db 19715 GATCAGCGCCGACATGCTGACCGGCTTCTGCGCTCTGCGCTGCGAGATGCTGCTGAGCG 19774
Qy 2006 aggaatgagcagtgatgactgactgactgactgactgactgactgactgactgactgact 2065
Db 19775 CGGGTTCGAGCGATGCTTCCAGATGGCCCGCTCGAAGCTGCGCAACCTGCGCAACGTC 19834
Qy 2066 atgattgatacactgtct 2083
Db 19835 ACCGTACGACAGCGCGT 19852

RESULT 2
LOCUS ECPOXB 1974 bp DNA BCT 12-SEP-1993
DEFINITION E. coli poxB gene for pyruvate oxidase.
ACCESSION X04105 M13947 M13948
VERSION X04105.1 GI:42472
FEATURES
flavoprotein; inverted repeat; poxB gene; pyruvate oxidase.

SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
1 (bases 1 to 1974)
REFERENCE Grabau, C. and Cronan, J.E. Jr.
AUTHORS Nucleotide sequence and deduced amino acid sequence of Escherichia
TITLE coli pyruvate oxidase, a lipid-activated flavoprotein
JOURNAL Nucleic Acids Res. 14 (13), 5449-5460 (1986)
MEDLINE 86286555
COMMENT Data kindly reviewed (31-MAY-1987) by C. Grabau.
FEATURES
Location/Qualifiers
source
1..1974
/organism="Escherichia coli"
/strain="K12"
/db_xref="taxon:562"
56..64
/note="inverted repeat A"
73..81
/note="inverted repeat A"
164..168
/note="put. ribosome binding site"
179..1897
/codon_start=1
/transl_table=11
/product="pyruvate oxidase"
/protein_id="CAA27725.1"
/db_xref="GI:42473"
/db_xref="SWISS-PROT:P07003"
/translation="MKQTVAAYIAKTLKESAGVKRIMGVTDGSLNGLSDSLNRMGTIEW
MSTRHEEVAFAAFAEAQALSGELAVCAAGCPGNLHNLGLFDCRHRNVPVLAFAAHI
PSSIGSGYFOETHPELFRECSHYCELVSPQIPOVLAIAMRKAVLNKGVVVLVLP
GDVALKPAPEGATMHVYHAPQVPTPEEELRKLAQLRYSNLNADLMLVLLGTQF
LYRFAGTKDAKIIQIDINPASIGAHKVDMAVGDIKSTLRALLPLVEEKADRRFLDK
ALEYDRARKGLDLAKPSEKAIHPOYLAQOISHPAADDALFTCDVGTPTVMAARYLK
MNGKRRLLGNSFNHSGMANNMPOALGAQATPEROVAVMGCGFSLMGDFLSVYQMK
LPVXIVFVNSVLGFVAMKRGYLTGTEHDNFIARIEAGGITGIRVEKASEVD
EALQARFSDIGPVLVDVVVAVKBEIAIPQIKLEQAKFSLYMLRALISGRGDEVIELA
KTNMLR"
BASE COUNT 474 a 521 c 535 g 444 t
ORIGIN
Query Match 14.1%; Score 305.2; DB 2; Length 1974;
Best Local Similarity 50.7%; Pred. No. 3e-70;
Matches 873; Conservative 0; Mismatches 828; Indels 21; Gaps 5;
Qy 327 atggcacacagctacgcagacaataattgacacatttggaaagctcaagtgtaagcga 386
Db 179 ATGAACAAACGGTTCGAGCTTATCGCCAAACACACTCGAATCGCGAGGGGTGAACGC 238
Qy 387 atttatggttgggtgagcagccttaa---tccgacgtggtgctgctccgccaatca 443
Db 239 ATCTGGGAGTACACAGCGACTCTCTGAACGGCTTTAGTCACACTTATATCGCATGGGC 298
Qy 444 gataatgagtggtgacgcttcgaaatgaggaagcggcgtttgagcgggtgagcga 503
Db 299 ACCATCGAGTGGATGTCACCCGCCACGAGAGAGTGGCGGCTTTGCCGCTGGCGCTGAA 358
Qy 504 tcggttgactcactgggagctggtgagtgatgctgctcttctggtctctggaacacacac 563
Db 359 GCACAACTTAGCGGAGAACTGGCGGTCTGCGCGGATGCTGCGGCCCGCGGAACCTGCAC 418
Qy 564 ctgattcaggtcttattgattcgcacgaaatggtcgaaaggtgttgccatcgctagc 623
Db 419 TTAATCAACGGCCTGTCGATGGCCACCGCAATCAGCTTCGGGTACTGCGCATTTGCCGCT 478
Qy 624 catattccagtgccagatggttgcagcgttcttccaggaaacagcatcccgagattttg 683
Db 479 CATATTCCCTCCAGCGAAATTTGGCAGCGGCTATTTCAGGAAACCCACCCACAGAGCTA 538
Qy 684 tttaaggaaatgctctggttactgagagatggtgaatggtgagcaggggtgaaacgcaatt 743